# Introducción a la minería de datos para la empresa

**Curso Académico: ( 2018 / 2019 )** **Fecha de revisión: 11-05-2018**

**Departamento asignado a la asignatura: Departamento de Estadística**

**Coordinador/a: MUÑOZ GARCIA, ALBERTO**

**Tipo: Optativa Créditos ECTS : 6.0**

**Curso : Cuatrimestre :**

## REQUISITOS (ASIGNATURAS O MATERIAS CUYO CONOCIMIENTO SE PRESUPONE)

Esta asignatura asume que el alumno conoce la materia impartida en:

a) Estadística I (http://www3.uc3m.es/reina/Fichas/Idioma_1/204.13154.html),
b) Estadística II (http://www3.uc3m.es/reina/Fichas/Idioma_1/204.13160.html),

y el tema de Propiedades de Matrices en
c) Matemáticas para la economía II (http://www3.uc3m.es/reina/Fichas/Idioma_1/204.13156.html)

del grado de Administración de Empresas.

## OBJETIVOS

COMPETENCIAS ESPECIFICAS
1. Conocer y utilizar técnicas estadísticas avanzadas, con soporte informático de última generación.
2. Extraer y analizar información de grandes conjuntos de datos.
3. Adquirir las habilidades estadísticas para el análisis de datos multivariantes socio-económicos, tales como los que aparecen en investigación de mercados.
4. Habilidad de describir y analizar conjuntos de datos reales con las técnicas mencionadas.
5. Capacidad de elaborar informes de resultados de los análisis sobre estudios de caso reales.

COMPETENCIAS TRANSVERSALES
1. Capacidad de análisis y síntesis de la información en problemas de Data Mining.
2. Resolución de problemas reales.
3. Conocimiento y adiestramiento en el uso de software estadístico para la resolución de problema

## DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

1. Aprendizaje del paquete estadístico R.
   1.1 Comandos básicos.
   1.2 Visualización de datos en R.
   1.3 Funciones de análisis de datos en R.
2. Técnicas de visualización para datos complejos de empresa y finanzas
   2.1 Teoría del análisis de componentes principales.
   2.2 Ejemplos básicos e implementación en R.
   2.3 Caso de estudio.
3. Distancias en Data Mining.
   3.1 Conceptos y distancias en data mining.
   3.2 Implementación de distancias en R.
1. Escalado multidimensional.
   4.1 Teoría del escalado métrico.
   4.2 Ejemplos en R.
   4.3 Mapas perceptuales en R.
2. Análisis de cluster (detección de grupos en datos complejos).
   5.1 Métodos jerárquicos.
   5.2 Métodos de centroides: k-medias.
   5.3 Casos de estudio.
3. Reglas de asociación.
   6.1 Principales conceptos.
   6.2 Ejemplo completo en R.
   6.3 Caso de estudio.

4. Árboles de clasificación.
   7.1 Teoría de la información.
   7.2 Algoritmos de construcción de árboles.
   7.3 Caso de estudio: credit scoring.
8. Casos de estudio reales adicionales.
8.1 Casos reales comprehensivos que utilizan todas las técnicas estudiadas.

## ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

1. Clases teóricas (4 créditos ECTS)
2. Clases prácticas con ordenador (2 créditos ECTS)
3. Trabajo final.

## SISTEMA DE EVALUACIÓN

50% de la calificación: Examen final.
10% de la calificación: Evaluación continua (*).
40% de la calificación: entrega del trabajo final.

(*) La evaluación continua consiste en la entrega de varias prácticas que se indicarán a lo largo del curso.

| | |
|---|---|
| **Peso porcentual del Examen Final:** | 60 |
| **Peso porcentual del resto de la evaluación:** | 40 |

## BIBLIOGRAFÍA BÁSICA

- Avril Coghlan A little book of R for multivariate analysis, Internet, 2014
- Johannes Ledolter Data Mining and Business Analytics with R, Wiley, 2013

## BIBLIOGRAFÍA COMPLEMENTARIA

- Y Zhao R and Data Mining. Examples and Case Studies, Elsevier, 2012